ASSOCIATIONS BETWEEN ALDH1A1 POLYMORPHISMS, ALCOHOL CONSUMPTION, AND MORTALITY AMONG HISPANIC AND NON-HISPANIC WHITE WOMEN DIAGNOSED WITH BREAST CANCER: THE BREAST CANCER HEALTH DISPARITIES STUDY

by

Zhiyu Xia

A master thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of

Master of Science

in Cancer Epidemiology, Department of Epidemiology

at Johns Hopkins Bloomberg School of Public Health

Baltimore, Maryland

April 21, 2017

ABSTRACT

Background

ALDH1A1 is a marker of both normal tissue and cancer stem cells, where it is involved in self-renewal, differentiation and self-protection. Few studies have examined the association between *ALDH1A1* and mortality among breast cancer (BC) patients. None of these studies have included Hispanic women or explored interactions with alcohol consumption. We evaluated the associations between *ALDH1A1* polymorphisms, alcohol consumption, and mortality among Hispanic and non-Hispanic white (NHW) BC cases from the Breast Cancer Health Disparities Study.

Methods

We evaluated the associations between nine SNPs of *ALDH1A1* and mortality among 920 Hispanic and 1372 NHW women diagnosed with incident invasive BC. Demographic and lifestyle factors were collected via in-person interviews. Additive, recessive, and dominant models were considered for each SNP. Cox proportional hazard regression models were used to calculate hazard ratios (HRs) and 95% confidence intervals (CIs), adjusting for age at diagnosis, percentage of Native American (NA) ancestry, BC stage, and study site. Models were further stratified by percentage of NA ancestry and alcohol consumption. Adjustment for multiple comparisons was performed using the Bonferroni correction.

Results

After a median follow-up time of 11 years from BC diagnosis, a total of 443 deaths occurred. The following SNPs were associated with increased all-cause mortality risk: the CC genotype of rs1424482 (HR_{CC}=1.31; 95% CI 1.03-1.68), the AA genotype of rs63319 (HR_{AA}=1.29; 95% CI 1.05-1.59), and the AA genotype of rs7027604 (HR_{AA}=1.40; 95% CI 1.13-1.73). rs722921 (TA/AA vs. TT) ($HR_{TA/AA}$=0.78; 95% CI 0.64-0.95) decreased risk of all-cause mortality. Only rs7027604 remained significant after adjustment for multiple comparisons (P_{adj}=0.018). Among ever drinkers, rs1888202 decreased mortality risk ($HR_{CG/GG}$=0.64; 95%CI 0.46-0.87), while no association was observed among non-drinkers ($P_{interaction}$=0.022, P_{adj}=0.181). Among women with low NA ancestry, rs63319 increased risk of mortality (HR_{AA}=1.53; 95%CI 1.19-1.97), while a non-significant inverse association was observed among women of high NA ancestry ($P_{interaction}$=0.022, P_{adj}=0.181). Results for BC-specific mortality were not statistically significant.

Conclusions

We provide evidence that rs7027604 is significantly associated increased risk of mortality after BC. Future BC studies examining the relationship between *ALDH1A1* and mortality should explore the modifying effects of alcohol consumption with rs1888202 and NA ancestry with rs63319.

PREFACE AND ACKNOWLEDGEMENTS

In August 2015, I began to pursue a Master of Science degree in cancer epidemiology, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health (JHSPH) in Baltimore, Maryland, USA. To fulfill the requirement of my degree, I worked on my master thesis as an independent researcher in Department of Epidemiology in JHSPH, advised by Dr. Avonne Connor.

My thesis project is focused on evaluating the underlying associations between ALDH1A1 polymorphisms, alcohol consumption, and mortality among Hispanic and non-Hispanic white women diagnosed with breast cancer. Few epidemiological studies have examined the association between *ALDH1A1* and mortality among women with breast cancer. To date, none of these studies have included Hispanic women or evaluated the modifying effect of alcohol consumption. To investigate my research question, I used data from the Breast Cancer Health Disparities Study (BCHDS), which is comprised of participants from three population-based case control studies. The genetic data was collected from DNA of either whole blood or mouthwash samples of the patients, and the survival data was obtained from cancer registries for each center. I conducted a time-to-event study with subgroup analyses.

ACKNOWLEDGEMENTS

This master thesis was made possible by the contributions of many individuals who have shared the journey with me. First and foremost, I want to express my deepest gratitude to my advisor Dr. Avonne Connor. I would like to thank you for all your tireless efforts and devotion, expertise, advice, and encouragement in helping me achieve the goal of finishing my master thesis, and also the consistent support and advice you provided me with during my two years in JHSPH. Acknowledgements are also due to Dr. Alison Klein, the reader of my thesis, for her meticulous and valuable comments on my thesis. I also would like to thank the research team of BCHDS, including all the investigators, researchers, and laboratory personnel, for their hard work and commitment to producing high quality clean data, which makes my study possible. I want to express my gratitude to JHSPH for the two years' graduate education: the systematic training in epidemiology I got here, the help from the supervisors of my RA positions, the support from the instructors of my TA job, and the encouragement from the most amazing cohort of fellow students, all of them have made valuable contributions to my academic and life experiences. Special thanks to my mom and dad, for their love, understanding and the appreciation they have instilled in me for seeking knowledge. I thank all the good friends by my side, for their company and encouragement along the way. There is no way I could be such a happy person without anyone of you. Thank you all!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

BACKGROUND

Breast cancer is one of the most common cancers in women aged 45-64 in the United States. According to the American Cancer Society, in 2015, approximately 231,840 females were diagnosed with breast cancer, among which about 40,290 died of the disease [1]. Breast cancer is a complex, multifactorial disease where there is a strong interplay between genetic (*BRCA1* and *BRCA2* [2]) and environmental factors (estrogen exposure [3], alcohol consumption [4], and dietary factors [5]). Furthermore, breast cancer is not a uniform disease, as in the U.S. alone, both incidence and mortality rates vary by time, geographic location, and by race and ethnicity.

1. Racial/ethnic disparities in breast cancer prognosis

1.1. The existence of racial/ethnic disparities

There are marked differences in the U.S. for breast cancer incidence and mortality rates by race and ethnicity [6]: breast cancer mortality is greater in African-American women compared with white American women, despite an inverse pattern reported for breast cancer incidence. Five-year breast carcinoma survival rates are 86% for white American patients compared to only 71% for African-American patients [7]. Even with the implementation of mammographic screening for early detection, breast carcinoma mortality rates continue to rise among African-American patients, while rates have begun to decline among white Americans [7]. Summary statistics from a meta-analysis of 14 studies revealed an odds ratio of 1.22 (95% confidence interval, 1.13–1.30) for the

adverse effect of African-American race on breast cancer mortality [8]. And the pooled all-cause mortality odds ratio was still significantly larger than 1 (1.27 with 95% confidence interval, 1.17-1.38) after adjusting for socioeconomic status (SES), indicating that African-American race was an independent predictor of mortality among women with breast cancer [8].

Similar results have been found among Hispanic women regarding breast cancer incidence and mortality rates [9-11]. Hispanic women also have higher mortality rates than non-Hispanic white (NHW) women, despite lower incidence rates [9, 10]. According to the results from a systematic review identifying several studies published between 1990 and 2007 [11], during the median follow-up period 6.3 years, the average incidence rates of Hispanic women was significantly lower than NHW women ($IRR_{Hispanic}$ 0.70, 95% confidence interval, 0.57-0.85) after adjusting for age. However, the estimated five-year breast cancer survival probability for Hispanic women was much lower than NHW women: the survival rate for Hispanics at 0-5 years was 82% versus 94% for NHWs, and the age-adjusted hazard ratios for Hispanic versus NHW was 3.4 (95% confidence interval, 1.75-6.61) [10].

Both Hispanic and African-American women are more likely to be diagnosed at later stages of breast cancer compared to white American women [12, 13]. According to a recent meta-analysis using data from 17 population-based cancer registries in the Surveillance, Epidemiology, and End Results (SEER) program [13], compared to NHW women, Hispanic women had 1.4-, 1.8-, 1.5- fold higher odds of presenting with stage II,

III, IV tumors respectively, meanwhile African-American women also had significantly higher odds of presenting with more advanced stages of breast carcinomas. African-American women along with Hispanic women had higher likelihood of being diagnosed with ER-/PR- breast cancer than NHWs (OR_{AA} 2.4, 95% confidence interval, 2.3-2.5; $OR_{Hispanic}$ 1.4, 95% confidence interval, 1.3-1.5); and among women with stage I or II breast carcinomas <2.0 cm in size, African-American women had an increased odds of receiving inappropriate primary surgical and radiation treatment (OR_{AA} 1.5, 95% confidence interval, 1.3-1.6), and a similar situation was also found among Hispanic women ($OR_{Hispanic}$ 1.2, 95% confidence interval, 1.1-1.3) [13]. In addition, both Hispanic and African-American female breast cancer patients tend to be younger [14]: among Hispanic and African-American patients, approximately 30% were diagnosed below age 50 years and 12% were diagnosed above age 75 years, however, those proportions were 18% and 21% respectively among NHW patients. Breast tumor sizes have also been reported to be different for Hispanic and African-American patients compared to NHW patients: Hispanic and African-American patients are more likely to have larger tumors (>3 cm) [10, 14].

1.2. Possible reasons for disparities

Causes of inequities in breast cancer survival between minority (including African-American and Hispanic) women and NHW women can be partially explained by the differences in the distribution of the modifiable and non-modifiable risk factors among these populations of women.

1.2.1. Clinical factors

Previous research has shown that clinical factors such as age at diagnosis, breast cancer stage, tumor size, estrogen and progesterone negativity, and comorbid medical conditions can largely explain the disparities in breast cancer mortality among African-American women and NHW women [15]. Since African-American patients tended to be younger, have larger tumor size, diagnosed at later stage, more likely to be ER-/PR-, and had more comorbidity, thus they experienced worse prognosis of breast cancer than NHW patients [15]. The relevant studies with U.S. Hispanic women are sparse, but since the distributions of those clinical factors are similar among African-American and Hispanic patients, and both of them had higher mortality rates than white Americans, we hypothesize clinical factors are essential to explain the ethnic disparities in mortality between Hispanic and NHW women.

1.2.2. Socioeconomic status (SES)

SES has been found to play a prominent role in explaining racial disparities in breast cancer mortality. However, the effect of SES on racial disparities in mortality may differ as the stage of breast cancer at diagnosis changes. According to Parise's research [16], for stage I patients, there was no significant difference between the risk of mortality among African-Americans and NHW within all SES strata; for stage II and stage III patients, African-Americans had the same risk of mortality as NHW in the lowest SES category but had an elevated risk of mortality in both intermediate and high SES groups; however, the hazard ratios for Hispanics versus NHW Americans were not significant regardless of disease stage and SES status [16]. Previous studies on how SES interplays with racial

disparities in breast cancer mortality have been inconsistent thus continued research is warranted.

1.2.3. Germline mutations distribution and genetic testing

Non-modifiable risk factors such as *BRCA1* mutations were found to be associated with higher risk for triple negative breast cancer [17]. A previous study demonstrated that *BRCA1* mutations were significantly more common in NHW women versus African-American women, while *BRCA2* mutations were more frequent in African-American women compared to NHW women [18]. Nevertheless, the frequency of *BRCA1/2* mutations were found to be dependent on the existence of family history of relevant cancers [19]. For those with family histories, NHW women had much higher prevalence of both *BRCA1* and *BRCA2* mutations than Hispanics and African-Americans; Hispanics had lower percentage of having *BRCA2* and similar proportion of having *BRCA1* compared to African-Americans. However, for those without family histories, Hispanic Americans had the highest prevalence of *BRCA1*, followed by white Americans, while African-Americans had the lowest prevalence. Even though the advancements in genetic testing would help reduce the breast cancer associated morbidity and mortality, research suggest that African American women are significantly less likely to receive genetic counseling and testing compared to white women [19]. The similar studies with Hispanic women are limited.

1.2.4. Screening and follow-up for abnormal screening tests

Access to breast cancer screening [20] and differences in quality of care among Hispanic, African-American and NHW women have also been found to contribute greatly to the disparities in breast cancer mortality [21]. Studies demonstrated significantly longer intervals for follow-up care after an abnormal mammogram for African-American women compared to white women, even after adjusting for insurance status [22].

1.2.5. Treatment

Differences in treatment preferences and quality of treatment are also factors to explain racial disparities in breast cancer mortality. According to the literature, both African-American and Hispanic patients were more likely to refuse having surgery and less likely to receive radiation therapy compared to the NHW patients [23]. Furthermore, only 69% of African-American patients received treatment within 30 days after diagnosis, while the proportion among NHW patients was 82% [22].

2. Breast cancer risk and prognostic factors

2.1. Age

Age is a well-documented risk factor affecting both incidence and mortality of breast cancer. According to Cancer Statistics 2015 [1], for women from age <50 years to ≥70 years, for each 10 years increased category, the probability of developing invasive breast cancer gradually increased from 1.9% to 12.3%. On the other hand, studies also demonstrated that women ≤40 years of age at diagnosis had significantly greater breast

cancer mortality compared to older populations even after controlling for socio-demographic factors, disease and treatment characteristics (HR 1.4, 95% confidence interval, 1.2-1.7) [24]. According to Partridge et al., the effect of age on survival of women with breast cancer was related to cancer subtypes: young age seems to be particularly prognostic in women with luminal breast cancers [24]. A Similar result was also found in another population-based study using the SEER 18 database [25], indicating that younger breast cancer patients exhibited more aggressive disease than older patients. The study also demonstrated that both young-aged (<40 years) and old-aged (≥60 years) patients had worse overall survival and breast cancer-specific survival compared to their middle-aged counterparts [25].

2.2. Reproductive factors

Reproductive factors including parity, breastfeeding, age at first birth, age at menarche, and age at menopause are found to be associated with breast cancer risk and prognosis. Parity and breastfeeding are well-documented protective factors for breast carcinomas [26-28]. According to the systematic review [26], nulliparous women were at elevated risk for breast cancer compared to parous women, with relative risk ranging between 1.2 and 1.7. The relationship between age at first birth and breast cancer risk is still controversial. Some studies found the older age at first birth, the higher risk of breast cancer [29], but the conclusions were not consistent. In a study conducted by Connor et al., researchers found that breastfeeding was a protective factor for both all-cause mortality and breast cancer-specific mortality among Hispanic and NHW women [27]. Kelsey's systematic review [26] also provided evidence that the younger a woman's age

at menarche and the later a women's age at menopause were both related to higher risk of breast cancer.

2.3. Obesity

Obesity is an important risk factor for developing postmenopausal breast cancer and also affects survival in women with breast cancer, but was found only among women that never used hormone replacement therapy (HRT) [30]. Among HRT non-users, heavier women (baseline BMI > 31.1) had an elevated risk of postmenopausal breast cancer (RR = 2.52; 95% CI 1.62–3.93) [30]. According to the meta-analysis conducted by Protani et al., which included 43 studies that enrolled women with breast cancer between 1963 and 2005, the pooled hazards of overall mortality among obese women with breast cancer was significantly higher than the hazards of women with normal BMI (HR 1.33, 95% confidence interval, 1.21-1.47) [31]. Similar results were also found for risk of breast cancer-specific mortality (HR 1.33, 95% confidence interval, 1.19-1.50) [31].

2.4. Family history

Family history of breast cancer has also been well documented as a significant predictor of both breast cancer risk and morality. In an analysis of data from 52 epidemiologic studies [32], breast cancer risk was significantly elevated for women having first-degree relatives with breast cancer, and the relative risk (RR) also increased dramatically with adding number of first-degree relatives having the disease. The results indicated the RR for having 1 first-degree relative with breast cancer was 2.14 (95% confidence interval, 1.92-2.38); for having 2 first-degree relative with breast cancer was 3.84 (95%

confidence interval, 2.37-6.22); and for having ≥3 first-degree relative with breast cancer was 12.05 (95% confidence interval, 1.70-85.16) [32]. Furthermore, the risk of breast cancer was higher among women with first-degree relatives who were diagnosed at younger ages compared to those diagnosed at older ages. Women with family history of breast cancer who were also diagnosed with breast cancer have also been found to have higher all-cause mortality than those without family history [32, 33].

2.5. Alcohol consumption

Alcohol consumption has been identified as a risk factor for breast cancer in many epidemiological studies [34]. There has also been strong evidence of a dose-response relationship between alcohol consumption and breast cancer risk [35]: there was a monotonic rise in the RR of breast cancer with increase alcohol consumption, however the magnitude of the effect was small. Women having approximately one typical drink consumption per day had a RR of 1.10 (95% confidence interval, 1.06-1.14) to develop breast cancer compared to those nondrinkers [35]. Even though alcohol intake has been consistently and positively associated with risk of breast cancer in women [34-39], no significant results were observed in the studies examining the relationship between alcohol consumption and overall mortality after breast cancer diagnosis [37], except for one study showing that women who consumed 2-3 drinks per day had a higher risk of breast cancer mortality (RR 1.5, 95% confidence interval, 1.2-1.9), however the mortality did not increase for those having at least 4 drinks a day [38]. The complicated risk-benefit ratio of alcohol intake on breast cancer mortality might due to the protective effect of

moderate alcohol consumption on cardiovascular disease and overall mortality [39]. More relevant studies are needed.

3. Aldehyde dehydrogenase 1A1 (ALDH1A1) gene

ALDH1A1 is a marker of normal tissue stem cells and cancer stem cells, where it is involved in self-renewal, differentiation and self-protection [40]. Studies have shown that high expression of ALDH1A1 was correlated with poor cancer prognosis, including ovarian cancer [41], prostate cancer [42] and lung cancer [43]. However, the results were not consistent in other cancers: the result indicated that high expression of ALDH1A1 was a prognostic marker for better survival in pancreatic cancer [44].

How ALDH1A1 expression affects the survival of patients with breast cancer remains unclear and contradictory. Some studies demonstrated that high ALDH1A1 expression was associated with later stage breast cancer, larger tumor size, chemo-resistance, prediction for early metastasis, thus having poorer prognosis [45-48]. Specific SNPs of ALDH1A1 including rs8187996, rs3764435 and rs63319 were found to be significantly associated with grade 3 and 4 hematological toxicity among patients treated with cyclophosphamide and doxorubicin for breast cancer, which might contribute to poorer prognosis of breast cancer as well [49]. However, other research indicated ALDH1A1 expression was independent to the prediction of breast cancer survival [50, 51], while one study reached an opposite conclusion reporting that high ALDH1A1 expression was associated with better clinical outcomes [52]. However, none of those studies included Hispanic women.

4. ALDH1A1 gene and alcohol consumption

ALDH1A1 is related to alcohol-induced flushing, alcohol sensitivity and dependence [53, 54], and functions in the detoxification of acetaldehyde, the first metabolite of ethanol oxidation [55]. The association between genetic variation in ALDH1A1 and the vulnerability to alcohol dependence and other alcohol-related phenotypes has been investigated in some studies. According to the sequencing results from Spence and colleagues [53], two rare promoter polymorphisms, ALDH1A1*2 and ALDH1A1*3 (which was only detected in African Americans) were sequenced. There was no association observed between ALDH1A1*2 and alcohol dependence, but there was a significant difference in ALDH1A1*3 allelic frequency between the alcoholics and the controls. Furthermore, the ALDH1A1*3 polymorphism resulted in reduced ALDH1A1 expression. ALDH1A1*2 was found to be associated with alcohol dependence according to some similar research among different study populations including African-Americans and American Indians [56-58], nevertheless, the evidence on populations Hispanic and NHW is very sparse.

5. Importance

Few epidemiological studies have examined the association between ALDH1A1 and mortality among women with breast cancer. To date, none of these studies have included Hispanic women or evaluated the modifying effect of alcohol consumption. These discrepancies and the scant body of literature in this research area highlight the need for more quantitative research to explore the underlying associations between ALDH1A1

polymorphisms, alcohol consumption, and mortality among Hispanic and NHW women diagnosed with breast cancer.

METHODS

Study design and population

The Breast Cancer Health Disparities Study is comprised of participants from three population-based case-control studies: the 4-Corners Breast Cancer Study (4-CBCS), the San Francisco Bay Area Breast Cancer Study (SFBCS), and the Mexico Breast Cancer Study (MBCS) [59]. For this particular study, the study participants from the MBCS and all population-based controls were excluded due to missing data for vital status. Institutional Review Board for Human Subjects was obtained at each institution and informed consent was acquired from all participants before participation. Each participant completed an interview and had a mouthwash or blood sample for DNA extraction to access the genetic data.

The 4-CBCS and the SFBCS have been described in detailed previously [59]. The 4-CBCS consisted of NHW, Hispanic and Native American (NA) women participants aged 25-79 years living in the areas of Arizona, Colorado, New Mexico, and Utah at the time of diagnosis or selection [60]. Cases were women histologically confirmed as either in situ or invasive breast cancer between October 1999 and May 2004. Controls were selected from the target populations and were frequently matched to cases on ethnicity and 5-year age distribution. A total of 2557 cases (1683 NHW, 852 Hispanics and 22 American Indian) and 2605 controls (1669 NHW, 913 Hispanics and 23 American Indian) were included into the 4-CBCS [59]. Respondents who reported being NA

precluded separate analysis for this group, thus NA women and Hispanic women were grouped together in all of the 4-CBCS analyses [61]. The SFBCS comprised of NHW and Hispanic women aged 35-79 years from the San Francisco Bay Area [59]. The cases were women diagnosed with confirmed breast cancer between April 1995 and April 2002; and the controls were selected by random-digit dialing and frequency matched based on ethnicity and 5-year age distribution of cases. A total of 1105 cases (312 NHW and 793 Hispanics) and 1318 controls (320 NHW and 998 Hispanics and) were included into the SFBCS [59].

For this study, the exclusion criteria were as follows: 1) all controls were excluded due to missing data pertaining to vital status and cause of death; 2) the cases that were not primary breast cancer; 3) the cases that were confirmed as in situ breast cancer; 4) the individuals whose cause of death was not specified; 5) had missing data for either percentage of NA ancestry or alcohol consumption. Our final study sample included a total of 2292 incident invasive breast cancer cases.

Data harmonization

Data were harmonized across all study centers and questionnaires [59]. The main variables for harmonization were selected based on study hypotheses and the genetic pathways of interest. The current study considered adjusting for age at breast cancer diagnosis, percentage of NA ancestry (categorized as low ≤ 0.28 versus high >0.28), breast cancer summary stage (localized, regional, distant, or unstaged), and study site.

Long-term alcohol consumption was defined as the average amount of alcohol consumption in grams over specific ages (i.e. 15, 30 and 50). For the present analysis, alcohol consumption was modeled as history of alcohol consumption vs. none.

Genetic data

DNA was extracted from either whole blood or mouthwash samples [62]. A total of 7286 blood-derived and 637 mouthwash-derived samples were available for DNA analyses. Whole genome amplification was applied to the mouthwash-derived DNA samples prior to genotyping. A tagSNP approach was used to characterize variation across candidate genes. TagSNPs were selected using the following parameters: linkage disequilibrium (LD) blocks were defined using a Caucasian LD map and an $r^2 = 0.8$; minor allele frequency >0.1; range = -1500 bps from the initiation codon to +1500 bps from the termination codon; and 1 SNP/LD bin [63]. 104 ancestral informative markers were used to distinguish European and NA ancestry in the study population. All markers were genotyped using a multiplexed bead array assay format based on GoldenGate chemistry (Illumina, San Diego, CA). A genotyping call rate of 99.93% was attained (99.65% for whole genome amplification samples). We included 132 internal replicates that were blinded representing 1.6% of the sample set. The duplicate concordance rate was 99.996% as determined by 193,297 matching genotypes among sample pairs [59].

In the present study, we examined nine SNPs in the ALDH1A1 gene (rs348481, rs168351, rs1888202, rs63319, rs722921, rs348461, rs348463, rs7027604, rs1424482).

The ALDH1A1 polymorphisms in detail, including the minor allele frequencies (MAFs) and adjusted Hardy-Weinberg equilibrium (HWE) p-values were available in the Table 1.

Survival data

Vital status was available for the Utah, New Mexico, Colorado, Arizona, and California study centers [63]. The cancer registry for each center provided information on the date of death or last date of follow-up [62]. Survival years were calculated as the date of death or last date of follow-up minus the date of diagnosis of breast cancer. For this analysis, causes of death were classified as breast cancer (death certificate noted breast cancer), other cancer (death certificate noted other types of cancer), and non-cancer.

Statistical methods

The program STRUCTURE was used to compute individual ancestry for study participants assuming two founding populations [64, 65]. The level of percentage NA ancestry was used to classify each participant into either low or high percentage of NA ancestry group (≤0.28 versus >0.28). The descriptive statistics were calculated for all covariates and t-tests and chi-square tests were used to assess the differences between groups. The homozygous common genotypes for each SNP were used as the referent categories. Modes of inheritance including additive, dominant and recessive models were considered for each SNP accordingly using their genotypes. Haploview was used to check the linkage disequilibrium (LD) between the nine SNPs of *ALDH1A1*. We found

two blocks on this gene: the SNP rs348461 and rs348463, and the SNP rs7027604 and rs722921 were in high LD, respectively. Since the effects of the SNP rs7027604 and rs722921 on breast cancer mortality were in opposite direction, and neither of the SNP rs348461 nor rs348463 were significantly associated with all-cause mortality, we decided to keep all nine SNPs and analysis them separately before adjusting for multiple comparisons. Multivariate Cox proportional hazard regression models were used to calculate hazard ratios (HRs) and 95% confidence intervals (CIs). Proportional hazard assumption was checked by using Stata command *estat phtest* for each SNP. The proportional hazard assumption was met for all models. We evaluated potential confounders including age at diagnosis, percentage of NA ancestry, study sites, breast cancer summary stage, menopausal status, use of oral contraceptives, use of HRT, alcohol consumption and parity by calculating the percentage change in HRs for all-cause mortality (significant confounding defined by change in main effect point estimate of more than 10%). After this evaluation, we found that none of these factors were significantly confounding the associations between the SNPs and mortality outcomes; therefore we adjusted the final models for age at diagnosis, percentage of NA ancestry, breast cancer summary stage, and study site to account for any residual confounding by these factors. To detect effect modification, interaction terms were created between percentage of NA ancestry (low $\leq$0.28 versus high >0.28), alcohol consumption (ever versus never) and the polymorphisms. The likelihood ratio test was used to evaluate whether the interaction terms were significant. The Bonferroni correction was used to adjust for multiple comparisons [66].

Power calculations were performed using the *PS: Power and Sample Size Calculation version 3.0 software* for the studies of survival analysis, which can be accessed at: http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize [67, 68]. We estimated the power of the analysis using the SNPs rs168351 (having the lowest minor allele frequency which equals to 0.1171), rs722921 (having the highest minor allele frequency which equals to 0.4953) and rs7027604 (minor allele frequency equals to 0.4527). The following parameters were considered to calculate the power (refer to Table 3 and Table 4): median survival time for patients in low NA ancestry group as 11 years, accrual time during which patients were recruited as 4 years, additional follow-up time after end of recruitment as 11 years, the number of patients who have minor allele as minor allele frequency of each SNP times total number of participants (0.1171*2292; 0.4953*2292; 0.4527*2292, respectively), ratio of patients having major allele versus patients having minor allele ($\frac{1-0.1171}{0.1171} = 7.53; \frac{1-0.4953}{0.4953} = 1.02;\ \frac{1-0.4527}{0.4527} = 1.22$) respectively, and defined type I error rate = 0.05, with HRs equal to 1.2, 1.5 and 2.0. Under the above conditions, the overall power ranged from 52.9% to 100.0% (Table 2).

All statistical analyses were performed using SAS (version 9.2), Haploview (version 4.2) and Stata (version 12.0).

RESULTS

Baseline characteristics

The baseline distributions of the genotype frequencies, demographics and major risk factors for breast cancer prognosis stratified by NA ancestry and alcohol consumption are shown in Table 3 and Table 4 respectively. A total of 2292 breast cancer cases were included in the present analysis, among which 1372 were in low NA ancestry category and 920 were in high NA ancestry category; the number of never-drinker was 1292 and the number of ever-drinker was 1000. After a median follow-up time of 11 years from breast cancer diagnosis, a total of 443 deaths occurred. The median survival time in low NA ancestry category was 10.83 years, which was statistically identical to the median survival time in high NA ancestry category (11 years).

For low NA ancestry group versus high NA ancestry group (≤0.28 vs. >0.28), the distributions of the genotype frequencies were significantly different for some polymorphisms including rs1424482, rs168351, rs348461, rs368463, rs348481 and rs7027604. The distributions of the major risk factors for breast cancer prognosis such as age at diagnosis, cancer stage, menopausal status, oral contraceptives usage, HRT, parity, alcohol consumption, smoking pack years, BMI and median survival time were also compared between low and high NA ancestry groups (Table 3). Compared to women in low NA ancestry category, women in high NA ancestry group tended to have smaller proportion of being localized cancer (57.0% versus 67.7%), larger proportion of being

pre/peri-menopausal status (41.0% versus 34.3%), fewer cases using oral contraceptives (59.1% versus 71.2%) and HRT (42.2% versus 64.4%), and women in high NA ancestry group had significantly larger number of parities than women in low NA ancestry category. In addition, low NA ancestry group had significantly larger proportion of women ever had alcohol consumption compared to high NA ancestry group (52.1% versus 31.2%). The women in low NA ancestry group tended to be diagnosed at significantly older age than women in high NA ancestry group (55.61 versus 53.57 years). The mean smoking pack year in low NA ancestry group was 8.73 with standard deviation 0.51, while the mean smoking pack year in high NA ancestry group was 2.97 with standard deviation 0.30: the difference between the two groups was tremendous. Women in low NA ancestry category had smaller mean BMI than women in high NA ancestry group (27.4 versus 29.5 kg/m^2).

For never-drinker versus ever-drinker, the distributions of the genotype frequencies were statistically identical for all nine SNPs (Table 4). Compared to the distributions of the major risk factors for breast cancer prognosis, oral contraceptive usage, HRT, parity, smoking pack years and BMI were associated with alcohol consumption. Ever-drinkers were more likely to use oral contraceptives and HRT than never-drinkers, and they also tended to be less parous. The mean smoking pack year for ever-drinkers was 9.96 with standard deviation 0.65, which was significantly higher than that for never-drinkers (3.91 with standard deviation 0.31), and ever-drinkers had lower average BMI than never-drinkers (27.1 versus 29.2 kg/m^2). The mean age was similar for never-drinkers and ever-drinkers (54.71 versus 54.90 years).

Multivariate Cox proportional hazard model for overall mortality and breast cancer-specific mortality

The results of the HRs, CIs and p-values from univariate and multivariate Cox proportional hazard models are shown in Table 5 and Table 8, for overall mortality and breast cancer-specific morality respectively. The following four SNPs were significantly associated with overall mortality: rs1424482 (HR_{CC} =1.31; 95% CI 1.03-1.68), rs63319 (HR_{AA} = 1.29; 95% CI 1.05-1.59), rs7027604 (HR_{AA} = 1.40; 95% CI 1.13-1.73), and rs722921 ($HR_{TA/AA}$ = 0.78; 95% CI 0.64-0.95). However, only rs7027604 remained significant after adjustment for multiple comparisons (P_{adj} = 0.018). Figure 1 shows the Kaplan-Meier survival curves for each genotype of the polymorphism rs7027604 taking into consideration the recessive mode of inheritance. The two survival curves for genotypes CC/CA and AA were significantly different (log rank p-value = 0.0043).

For breast cancer-specific morality, the four SNPs that were significantly associated with overall mortality were also related to breast cancer-specific morality with the same directional point estimates: rs1424482 (HR_{CC} =1.09; 95% CI 0.76-1.55), rs63319 (HR_{AA} = 1.15; 95% CI 0.86-1.53), rs7027604 (HR_{AA} = 1.21; 95% CI 0.90-1.63), and rs722921 ($HR_{TA/AA}$ = 0.89; 95% CI 0.67-1.17). However, the hazard ratios were not statistically significant.

Associations between ALDH1A1 polymorphisms, all-cause and breast cancer-specific mortality, stratified by alcohol consumption/NA ancestry

The results of the stratification analyses and multivariate Cox proportional hazard regression models that included interaction terms created between alcohol consumption (ever versus never), level of NA ancestry (low versus high) and the SNPs for overall mortality are shown in Table 6 and Table 7, respectively. Among women with low percentage of NA ancestry (majority NHW women), rs63319 significantly increased risk of overall mortality (HR_{AA} = 1.53; 95% CI 1.19-1.97), while a non-significant inverse association was observed among women of high percentage of NA ancestry (mostly Hispanic women) (HR_{AA} = 0.92; 95% CI 0.64-1.34). Among ever drinkers, rs1888202 significantly decreased risk of all-cause mortality ($HR_{CG/GG}$ = 0.64; 95%CI 0.46-0.87), while no association was observed among non-drinkers ($HR_{CG/GG}$ = 1.04; 95%CI 0.79-1.36). However, the interaction terms for these two SNPs were not significant when conditioned on multiple comparisons. The results of the stratification analyses and multivariate Cox proportional hazard regression models that included interaction terms created between alcohol consumption (ever versus never), level of NA ancestry (low versus high) and the SNPs for breast cancer-specific mortality are shown in Table 9 and Table 10. Though the point estimates of hazard ratios were in the opposite direction between the different NA ancestry categories or alcohol consumption groups, we didn't reach statistical significance.

DISCUSSION

In this study that consisted of a large sample of Hispanic and NHW women from the Breast Cancer Health Disparities Study, we evaluated the associations of 9 polymorphisms of ALDH1A1 with breast cancer-specific mortality and overall mortality after invasive breast cancer diagnosis. We observed that some polymorphisms of ALDH1A1, including rs1424482, rs63319, rs7027604 and rs722921 were significantly associated with worse breast cancer prognosis, conditioning on age at diagnosis, level of NA ancestry, disease stage, and study site. Taking into account multiple comparisons, rs7027604 was still statistically associated with worse breast cancer prognosis. When stratified by level of percentage NA ancestry (low versus high), we found rs63319 significantly increased the risk of overall mortality among women with low percentage of NA ancestry, while a non-significant decrease in risk of all-cause mortality was observed among women in the high NA ancestry category. When stratified by alcohol consumption (ever versus never), rs1888202 significantly decreased the risk of mortality among ever drinkers, however no protective association was observed among never drinkers. If longer follow-up time permitted and more breast cancer cases were captured, we would expect the results to be similar for breast cancer-specific mortality. These findings support our hypotheses that polymorphisms from the ALDH1A1 gene are related to breast cancer prognosis, and these associations may be modified by level of NA ancestry and alcohol consumption.

Consistent with our findings, Ginestier et al. reported that ALDH1 is a marker of stem/progenitor cells of the normal human breast and breast carcinomas [45], and ALDH1-positive tumors were strongly associated with poorer clinical outcome compared to ALDH1-negative tumors: the 5 year overall survival was 69.59% (95% CI 60.73 – 79.73) for patients with an ALDH1-positive tumors and 84.55% (95% CI 80.02 – 89.33) for patients with ALDH-negative tumors [45]. Zhong et al. studied the expression of ALDH1A1 in invasive breast carcinoma, where they pointed out that ALDH1A1 expression was an independent predictor for recurrence-free survival and distant metastasis-free survival of breast cancer, and higher ALDH1A1 expression was associated with worse breast cancer prognosis [46]. The potential reasons to explain the association between higher ALDH1A1 expression and worse cancer prognosis include that the proportion of ALDH1-positive tumor cells together with CD44+/CD24- phenotypes were found to be significantly high when recurrence and metastasis occurred ($P = 0.019$) [47]; a positive relationship has also been found between ALDH1 expression and Ki-67 (a nuclear protein that is necessary for cellular proliferation) in invasive breast ductal carcinoma [46]; and that the metastasis, aggressive behaviors of breast cancer may be mediated by a cancer stem cells component that displays ALDH enzymatic activity [48].

The modifying effect of genetic ancestry on the association between ALDH1A1 polymorphisms and breast cancer prognosis has not been studied, to date. There is one report illustrating how expression of ALDH1 acted as a marker to predict breast cancer prognosis in African women [69]. According to Schwartz et al., among African women,

of the ER-, PR-, and HER2- defined subtypes of breast cancer, expression of ALDH1 was highest in triple-negative breast cancer, which has been found to be more aggressive and thus has a poorer prognosis compared to other subtypes of breast cancer due to its treatment complexities [69]. In our study, rs63319 significantly increased the risk of overall mortality among women with low percentage of NA ancestry (mostly NHW), while a non-statistically significant inverse association was observed with risk of all-cause mortality among women of high percentage of NA ancestry (mostly self-reported Hispanics). The reasons to explain the modifying effect of NA ancestry percentage on the associations between ALDH1A1 polymorphisms and breast cancer prognosis could be that NHW women and Hispanic women are exposed to different mutagens or endogenous factors, resulting in different expression levels of the ALDH1A1 gene. It is also possible that the existence of other unmeasured or unidentified genetic variants in or near the ALDH1A1 gene could make a difference in our findings. Different allele frequencies in the promoter regions between two populations might indicate different transcriptional activities, which could influence the ethnic differences observed in our study. However, more relevant studies at both population level and molecular level are required in the future to reveal the underlying mechanism of how percentage of NA ancestry modifies the association between ALDH1A1 and all-cause mortality among breast cancer patients. Another study conducted by Liu et al. analyzing four independent populations illustrated that the associations between haplotype blocks across ALDH1A1 and alcohol dependence were significantly different among plains Indians, American Indians, Finnish Caucasians and African Americans [70]. If we take into consideration the study results by Liu et al. and our results, it could be inferred that there might be three way interactions

between ALDH1A1 polymorphisms, population stratification and alcohol dependence, providing directions for relevant studies in the future.

To date, there are two papers that have demonstrated how ALDH1A1 is associated with alcohol-induced flushing, alcohol sensitivity and dependence [53, 54]. According to the sequencing results from Spence and colleagues [53], ALDH1A1*3 (one type of promoter polymorphism) was observed to have higher frequencies in the population of alcoholics, and ALDH1A1*3 expression was also found to be related to reduced ALDH1A1 expression [53]. In combination with the literature mentioned previously [45-48], we hypothesized that alcohol consumption might be a protective factor against mortality for women with breast cancer, which was proven to be plausible according to our results. We found that among ever drinkers, rs1888202 significantly decreased risk of mortality ($HR_{CG/GG}$ = 0.64; 95%CI 0.46-0.87), while no association ($HR_{CG/GG}$ = 1.04; 95%CI 0.79-1.36) was observed among non-drinkers. Consequently, alcohol consumption might have some modifying effect on the association between ALDH1A1 and mortality among women with breast cancer.

The present analysis has several strengths and some limitations. Our study is one of the very few epidemiologic studies examining the association between ALDH1A1 and mortality among women with breast cancer within a diverse study population, including Hispanic and NHW women from various geographical areas. And to date, none of the studies have evaluated the modifying effect of either alcohol consumption and percentage of NA ancestry on this association. We not only compared the outcome of all-cause

mortality but also analyzed breast cancer-specific mortality. However, our study was limited by the sample size to consider breast cancer-specific mortality, as we were unable to capture more deaths in this study, resulting in lack of power in both multiple comparisons and breast cancer-specific mortality analyses. Additional follow-up time and larger sample size would be strengths for the breast cancer mortality analyses. We had missing values for the subtypes of breast cancer and breast cancer treatment, thus subtypes and treatment were not included into the models. However, we adjusted for breast cancer stage in the multivariate Cox proportional hazard regression models, which is a strong predictor for mortality. In addition, we did not model the recurrence of breast cancer in the analyses; the results might be enhanced or different if we include recurrent breast cancer cases into the model. Furthermore, the AIMs we used to control for population substructure may not be sufficient enough to capture all the ancestry data thus having residual confounding, nevertheless, the Breast Cancer Health Disparities Study (BCHDS) was designed and powered to address differences in breast cancer risk and prognosis based on these AIMs. The stratification analyses and interactions should be examined further in future studies with larger sample sizes. The association between *ALDH1A1* and breast cancer-specific mortality should also be investigated with larger size of study populations and with more events of breast cancer-specific deaths in the future.

In summary, our study provides evidence that rs7027604 is associated with worse prognosis after breast cancer diagnosis among Hispanic and NHW women. Future breast cancer survival studies examining the relationship between *ALDH1A1* and mortality

should also explore the modifying effects of genetic ancestry with rs63319 and alcohol consumption with rs1888202.

TABLES

Table 1. Description of *ALDH1A1* polymorphisms by ethnicity, The Breast Cancer Health Disparities Study

SNP ID	Chromo-some Location	Coordinate	Region	Major/ minor Allele*	Non-Hispanic White (NHW)		Hispanic		Proportion missing
					Minor allele frequency	FDR adjusted HWE p-value	Minor allele frequency	FDR adjusted HWE p-value	
rs1424482	9q21.13	75563557	INTRON	T/C	0.35	0.96	0.43	0.63	0
rs168351	9q21.13	75517311	INTRON	A/G	0.17	0.86	0.07	0.12	0
rs1888202	9q21.13	75519251	INTRON	C/G	0.49	0.93	0.49	0.70	0
rs348461	9q21.13	75545070	INTRON	T/A	0.37	0.68	0.41	0.42	0.0005
rs348463	9q21.13	75547612	INTRON	T/C	0.28	0.96	0.39	0.39	0.0005
rs348481	9q21.13	75514436	INTERGENIC	T/C	0.23	0.49	0.12	0.10	0
rs63319	9q21.13	75524784	INTRON	C/A	0.49	0.87	0.48	0.46	0.0005
rs7027604	9q21.13	75554952	INTRON	C/A	0.42	0.96	0.48	0.82	0
rs722921	9q21.13	75544299	INTRON	T/A	0.49	0.96	0.50	0.95	0

*Major/minor allele reported for NHW population; minor allele frequency and Hard-Weinberg Equilibrium (HWE) based on control population.

Table 2. Power calculation using minor allele frequencies of the *ALDH1A1* SNPs rs168351, rs722921 and rs7027604, The Breast Cancer Health Disparities Study

	Hazard Ratios		
	1.2	1.5	2.0
rs168351 (A/G)*			
Minor allele frequency = 0.1171	52.9%	99.1%	100.0%
rs722921 (T/A)			
Minor allele frequency = 0.4953	88.5%	100.0%	100.0%
rs7027604 (C/A)			
Minor allele frequency = 0.4527	88.2%	100.0%	100.0%

* Major/minor allele

Table 3. Characteristics of study population, stratified by Native American (NA) ancestry, The Breast Cancer Health Disparities Study (N = 2292)

SNPs	Percentage (%) of NA Ancestry			P-value
	Overall (N=2292)	% NA Ancestry≤0.28¶ (n=1372)	% NA Ancestry>0.28 (n=920)	
rs1424482				**<0.001***
T/T	878 (38.3)	569 (41.5)	309 (33.6)	
T/C	1072 (46.8)	627 (45.7)	445 (48.4)	
C/C	342 (14.9)	176 (12.8)	166 (18.0)	
rs168351				**<0.001***
A/A	1757 (76.7)	960 (70.0)	797 (86.6)	
A/G	493 (21.5)	372 (27.1)	121 (13.2)	
G/G	42 (1.8)	40 (2.9)	2 (0.2)	
rs1888202				0.880
C/C	588 (25.6)	357 (26.0)	231 (25.1)	
C/G	1,157 (50.5)	688 (50.2)	469 (50.9)	
G/G	547 (23.9)	327 (23.8)	220 (24.0)	
rs348461				**0.042***
T/T	850 (37.1)	528 (38.5)	322 (35.0)	
T/A	1085 (47.3)	650 (47.4)	435 (47.3)	
A/A	357 (15.6)	194 (14.1)	163 (17.7)	
rs348463				**<0.001***
T/T	1074 (46.8)	700 (51.0)	374 (40.7)	
T/C	978 (42.7)	565 (41.2)	413 (44.9)	
C/C	240 (10.5)	107 (7.8)	133 (14.4)	
rs348481				**<0.001***
T/T	1525 (66.5)	815 (59.4)	710 (77.2)	
T/C	684 (29.9)	484 (35.3)	200 (21.7)	
C/C	83 (3.6)	73 (5.3)	10 (1.1)	
rs63319				0.089
C/C	571 (24.9)	331 (24.1)	240 (26.1)	
C/A	1135 (49.5)	668 (48.7)	467 (50.8)	
A/A	586 (25.6)	373 (27.2)	213 (23.1)	
rs7027604				**0.032***
C/C	719 (31.4)	456 (33.2)	263 (28.6)	
C/A	1100 (48.0)	651 (47.5)	449 (48.8)	
A/A	473 (20.6)	265 (19.3)	208 (22.6)	
rs722921				0.383
T/T	606 (26.4)	371 (27.0)	235 (25.6)	
T/A	1113 (48.6)	650 (47.4)	463 (50.3)	
A/A	573 (25.0)	351 (25.6)	222 (24.1)	
Study Sites				**<0.001***

4-Corners	1353 (59.0)	974 (71.0)	379 (41.2)	
SFBCS	939 (41.0)	398 (29.0)	541 (58.8)	
Stage of cancer				**<0.001***
Localized	1453 (63.4)	929 (67.7)	524 (57.0)	
Regional	732 (31.9)	398 (29.0)	334 (36.3)	
Distant	23 (1.0)	15 (1.1)	8 (0.9)	
Unstaged	84 (3.7)	30 (2.2)	54 (5.9)	
Menopause				**0.002***
Pre/Peri menopause	817 (37.0)	458 (34.3)	359 (41.0)	
Post menopause	1393 (63.0)	876 (65.7)	517 (59.0)	
Oral contraceptives				**<0.001***
Ever	1505 (66.3)	967 (71.2)	538 (59.1)	
Never	765 (33.7)	392 (28.8)	373 (40.9)	
HRT				**<0.001***
Ever	1076 (55.2)	736 (64.4)	340 (42.2)	
Never	872 (44.8)	407 (35.6)	465 (57.8)	
Parity				**<0.001***
Nulliparous	330 (14.4)	232 (16.9)	98 (10.65)	
1-2	953 (41.6)	616 (44.9)	337 (36.6)	
3-4	764 (33.3)	424 (30.9)	340 (37.0)	
≥5	245 (10.7)	100 (7.3)	145 (15.75)	
Alcohol consumption				**<0.001***
Ever	1002 (43.7)	715 (52.1)	287 (31.2)	
Never	1292 (56.3)	658 (47.9)	634 (68.8)	
Age at diagnosis				**<0.001***
Mean±SD	54.79±11.03	55.61±0.30	53.57±0.36	
Smoking (pack yrs)				**<0.001***
Mean±SD	6.40±14.01	8.73±0.51	2.97±0.30	
BMI (Kg/m^2)				**<0.001***
Mean±SD	28.3±6.06	27.4±0.16	29.5±0.20	
Survival time				0.600
Median	10.92	10.83	11	

¶ Percent of genetic admixture from NA ancestry (0 would be no NA ancestry and 1 would be only NA ancestry)

Table 4. Characteristics of study population, stratified by alcohol consumption, The Breast Cancer Health Disparities Study (N = 2292)

SNPs/Covariates	Alcohol consumption			P-value
	Overall (N = 2292)	Never (n = 1292)	Ever (n = 1000)	
rs1424482				0.529
T/T	878 (38.3)	482 (37.3)	396 (39.6)	
T/C	1072 (46.8)	613 (47.45)	459 (45.9)	
C/C	342 (14.9)	197 (15.25)	145 (14.5)	
rs168351				0.075
A/A	1757 (76.7)	1011 (78.2)	746 (74.6)	
A/G	493 (21.5)	262 (20.3)	231 (23.1)	
G/G	42 (1.8)	19 (1.5)	23 (2.3)	
rs1888202				0.306
C/C	588 (25.6)	347 (26.9)	241 (24.1)	
C/G	1,157 (50.5)	645 (49.9)	512 (51.2)	
G/G	547 (23.9)	300 (23.2)	247 (24.7)	
rs348461				0.354
T/T	850 (37.1)	469 (36.3)	381 (38.1)	
T/A	1085 (47.3)	610 (47.2)	475 (47.5)	
A/A	357 (15.6)	213 (16.5)	144 (14.4)	
rs348463				0.158
T/T	1074 (46.9)	583 (45.1)	491 (49.1)	
T/C	978 (42.7)	567 (43.9)	411 (41.1)	
C/C	240 (10.5)	142 (11.0)	98 (9.8)	
rs348481				0.064
T/T	1525 (66.6)	875 (67.7)	650 (65.0)	
T/C	684 (29.8)	380 (29.4)	304 (30.4)	
C/C	83 (3.6)	37 (2.9)	46 (4.6)	
rs63319				0.730
C/C	571 (24.9)	314 (24.3)	257 (24.7)	
C/A	1135 (49.5)	647 (50.1)	488 (48.8)	
A/A	586 (25.6)	331 (25.6)	255 (25.5)	
rs7027604				0.466
C/C	719 (31.4)	392 (30.4)	327 (32.7)	
C/A	1100 (48.0)	627 (48.5)	473 (47.3)	
A/A	473 (20.6)	273 (21.1)	200 (20.0)	
rs722921				0.591
T/T	606 (26.4)	352 (27.25)	254 (25.4)	
T/A	1113 (48.6)	623 (48.2)	490 (49.0)	
A/A	573 (25.0)	317 (24.55)	256 (25.6)	
Study Sites				0.208
4-Corners	1353 (59.0)	748 (57.9)	605 (60.5)	

SFBCS	939 (41.0)	544 (42.1)	395 (39.5)	
Stage of cancer				0.069
Localized	1453 (63.4)	797 (61.7)	656 (65.6)	
Regional	732 (31.9)	424 (32.8)	308 (30.8)	
Distant	23 (1.0)	17 (1.3)	6 (0.6)	
Unstaged	84 (3.7)	54 (4.2)	30 (3.0)	
Menopause				0.729
Pre/Peri menopause	817 (37.0)	466 (37.3)	351 (36.6)	
Post menopause	1393 (63.0)	784 (62.7)	609 (63.4)	
Oral contraceptives				**<0.001***
Ever	1505 (66.3)	780 (61.0)	725 (73.1)	
Never	765 (33.7)	498 (39.0)	267 (26.9)	
HRT				**<0.001***
Ever	1076 (55.2)	553 (50.6)	523 (61.1)	
Never	872 (44.8)	539 (49.4)	333 (38.9)	
Parity				**<0.001***
Nulliparous	330 (14.4)	148 (11.5)	182 (18.2)	
1-2	953 (41.6)	462 (35.8)	491 (49.1)	
3-4	764 (33.3)	494 (38.2)	270 (27.0)	
≥5	245 (10.7)	188 (14.5)	57 (5.7)	
Age at diagnosis				0.683
Mean±SD	54.79±11.03	54.71±0.32	54.90±0.34	
Smoking (pack yrs)				**<0.001***
Mean±SD	6.40±14.01	3.91±0.31	9.96±0.65	
BMI (Kg/m^2)				**<0.001***
Mean±SD	28.3±6.06	29.2±0.17	27.1±0.18	

Table 5. Univariate and multivariable Cox proportional hazard models for all-cause mortality, The Breast Cancer Health Disparities Study

SNPs	N	No. death	HR§	P-value	HR¶	P-value	P_{adj} £
rs1424482							
T/T or T/C	1950	364	1.00		1.00		
C/C	342	79	1.25 (0.98-1.60)	0.068	1.31 (1.03-1.68)	**0.029***	0.233
rs168351							
A/A	1757	345	1.00		1.00		
A/G	493	86	0.88 (0.69-1.11)	0.289	0.93 (0.73-1.19)	0.572	0.999
G/G	42	12	1.49 (0.84-2.66)	0.172	1.51 (0.84-2.70)	0.168	0.809
rs1888202							
C/C	588	128	1.00		1.00		
C/G or G/G	1704	315	0.82 (0.67-1.00)	0.055	0.85 (0.69-1.04)	0.113	0.660
rs348461							
T/T	850	151	1.00		1.00		
T/A	1085	220	1.19 (0.97-1.46)	0.100	1.21 (0.98-1.49)	0.069	0.475
A/A	357	72	1.15 (0.87-1.52)	0.330	1.24 (0.94-1.65)	0.129	0.712
rs348463							
T/T	1074	197	1.00		1.00		
T/C	978	196	1.12 (0.92-1.37)	0.252	1.14 (0.93-1.39)	0.200	0.866
C/C	240	50	1.14 (0.83-1.56)	0.397	1.24 (0.91-1.70)	0.175	0.823
rs348481							
T/T	1525	302	1.00		1.00		
T/C	684	126	0.94 (0.76-1.15)	0.545	0.90 (0.73-1.11)	0.321	0.969
C/C	83	15	0.91 (0.54-1.53)	0.718	0.91 (0.54-1.53)	0.717	0.999
rs63319							
C/C or C/A	1706	314	1.00		1.00		
A/A	586	129	1.23 (1.00-1.51)	**0.047***	1.29 (1.05-1.59)	**0.015***	0.127
rs7027604							
C/C or C/A	1819	329	1.00		1.00		
A/A	473	114	1.36 (1.10-1.69)	**0.004***	1.40 (1.13-1.73)	**0.002***	**0.018***
rs722921							
T/T	606	137	1.00		1.00		
T/A or A/A	1686	306	0.79 (0.64-0.96)	**0.020***	0.78 (0.64-0.95)	**0.015***	0.127

§ Crude Cox Proportional Hazard model with only SNPs as covariate

¶ Adjusted- Cox Proportional Hazard model adjusting for age at diagnosis, percentage of Native American (NA) ancestry, disease stage, and study site.

£ Adjusted p-value using multiple comparison Bonferroni correction.

Table 6. Associations between *ALDH1A1* polymorphisms and all-cause mortality, stratified by alcohol consumption, The Breast Cancer Health Disparities Study

	Alcohol consumption						$P_{interaction}$	P_{adj}
	Never			Ever				
	HR	**95% CI**	**P-value**	**HR**	**95% CI**	**P-value**		
rs1424482							0.779	0.999
T/T or T/C	1.00	-		1.00	-			
C/C	1.34	0.98 - 1.85	0.070	1.29	0.88 – 1.90	0.198		
rs168351							0.437	0.994
A/A	1.00	-		1.00	-			
A/G	1.05	0.77 - 1.45	0.737	0.80	0.55 - 1.16	0.210		
G/G	1.59	0.70 - 3.63	0.268	1.40	0.61 - 3.19	0.427		
rs1888202							**0.022***	0.181
C/C	1.00	-		1.00	-			
C/G or G/G	1.04	0.79 - 1.36	0.801	0.64	0.46 - 0.87	**0.005***		
rs348461							0.723	0.999
T/T	1.00	-		1.00	-			
T/A	1.28	0.98- 1.68	0.071	1.11	0.80 - 1.53	0.549		
A/A	1.25	0.86 - 1.80	0.238	1.20	0.77 - 1.88	0.411		
rs348463							0.547	0.999
T/T	1.00	-		1.00	-			
T/C	1.27	0.98 - 1.64	0.071	0.94	0.68 - 1.29	0.699		
C/C	1.20	0.79 - 1.82	0.402	1.30	0.81 - 2.09	0.278		
rs348481							0.481	0.997
T/T	1.00	-		1.00	-			
T/C	0.93	0.71 - 1.23	0.615	0.86	0.62 - 1.19	0.365		
C/C	1.21	0.62 - 2.40	0.576	0.67	0.29 - 1.52	0.336		
rs63319							0.173	0.819
C/C or C/A	1.00	-		1.00	-			
A/A	1.14	0.86 - 1.50	0.364	1.53	1.12 - 2.09	**0.008***		
rs7027604							0.318	0.968
C/C or C/A	1.00	-		1.00	-			
A/A	1.27	0.96 - 1.68	0.094	1.62	1.17 - 2.26	**0.004***		
rs722921							0.356	0.981
T/T	1.00	-		1.00	-			
T/A or A/A	0.85	0.65 - 1.10	0.215	0.69	0.51 - 0.95	**0.023***		

Table 7. Associations between *ALDH1A1* polymorphisms and all-cause mortality, stratified by Native American (NA) ancestry, The Breast Cancer Health Disparities Study

	Percentage (%) of NA Ancestry						**$P_{interaction}$**	**P_{adj}**
	% NA Ancestry ≤ 0.28			**% NA Ancestry > 0.28**				
	HR	**95% CI**	**P-value**	**HR**	**95% CI**	**P-value**		
rs1424482							0.642	0.999
T/T or T/C	1.00	-		1.00	-			
C/C	1.40	1.01 – 1.93	**0.041***	1.23	0.84 - 1.79	0.292		
rs168351							0.494	0.999
A/A	1.00	-		1.00	-			
A/G	0.95	0.72 - 1.26	0.745	0.90	0.57 - 1.45	0.676		
G/G	1.55	0.86 - 2.78	0.143	-	-	-		
rs1888202							0.512	0.999
C/C	1.00	-		1.00	-			
C/G or G/G	0.81	0.63-1.06	0.121	0.92	0.65 - 1.30	0.636		
rs348461							0.378	0.986
T/T	1.00	-		1.00	-			
T/A	1.28	0.98 - 1.67	0.069	1.12	0.80 - 1.57	0.498		
A/A	1.39	0.97 - 2.00	0.076	1.08	0.69 - 1.69	0.732		
rs348463							0.318	0.968
T/T	1.00	-		1.00	-			
T/C	1.17	0.91 - 1.51	0.209	1.10	0.79 - 1.52	0.584		
C/C	1.50	0.98 - 2.30	0.061	1.04	0.66 - 1.67	0.854		
rs348481							0.146	0.758
T/T	1.00	-		1.00	-			
T/C	0.99	0.77-1.28	0.951	0.74	0.50 - 1.10	0.135		
C/C	0.99	0.57-1.71	0.975	0.54	0.07 - 3.87	0.537		
rs63319							**0.022***	0.181
C/C or C/A	1.00	-		1.00	-			
A/A	1.53	1.19 - 1.97	**0.001***	0.92	0.64 - 1.34	0.676		
rs7027604							0.205	0.873
C/C or C/A	1.00	-		1.00	-			
A/A	1.56	1.19 - 2.04	**0.001***	1.18	0.83 - 1.67	0.357		
rs722921							0.397	0.990
T/T	1.00	-		1.00	-			
T/A or A/A	0.73	0.57 - 0.95	**0.017***	0.87	0.62 - 1.22	0.420		

Table 8. Univariate and multivariate Cox proportional hazard models for breast cancer-specific mortality, The Breast Cancer Health Disparities Study

SNPs	N	No. death	HR§	P-value	HR¶	P-value
rs1424482						
T/T or T/C	1950	211	1.00		1.0	
C/C	342	36	0.98 (0.69-1.40)	0.923	1.09 (0.76-1.55)	0.648
rs168351						
A/A	1757	198	1.00		1.00	
A/G	493	43	0.77 (0.55-1.06)	0.111	0.87 (0.62-1.22)	0.414
G/G	42	6	1.30 (0.58-2.92)	0.530	1.66 (0.73-3.78)	0.229
rs1888202						
C/C	588	65	1.00		1.00	
C/G or G/G	1704	182	0.94 (0.71-1.24)	0.646	0.95 (0.71-1.26)	0.707
rs348461						
T/T	850	94	1.00		1.00	
T/A	1085	119	1.03 (0.78-1.35)	0.849	1.01 (0.77-1.32)	0.960
A/A	357	34	0.87 (0.59-1.29)	0.487	0.97 (0.65-1.44)	0.875
rs348463						
T/T	1074	112	1.00		1.00	
T/C	978	111	1.12 (0.86-1.45)	0.415	1.10 (0.84-1.43)	0.496
C/C	240	24	0.96 (0.62-1.50)	0.868	1.03 (0.66-1.61)	0.903
rs348481						
T/T	1525	173	1.00		1.00	
T/C	684	68	0.88 (0.66-1.16)	0.365	0.90 (0.68-1.19)	0.460
C/C	83	6	0.63 (0.28-1.43)	0.273	0.73 (0.32-1.65)	0.445
rs63319						
C/C or C/A	1706	184	1.00		1.00	
A/A	586	63	1.02 (0.77-1.36)	0.897	1.15 (0.86-1.53)	0.342
rs7027604						
C/C or C/A	1819	189	1.00		1.00	
A/A	473	58	1.20 (0.89-1.61)	0.226	1.21 (0.90-1.63)	0.202
rs722921						
T/T	606	69	1.00		1.00	
T/A or A/A	1686	178	0.91 (0.69-1.21)	0.523	0.89 (0.67-1.17)	0.403

§ Crude Cox Proportional Hazard model with only SNPs as covariate

¶ Adjusted- Cox Proportional Hazard model adjusting for age at diagnosis, percentage of Native American (NA) ancestry, disease stage, and study site.

£ Adjusted p-value using multiple comparison Bonferroni correction.

Table 9. Associations between *ALDH1A1* polymorphisms and breast cancer-specific mortality, stratified by alcohol consumption, The Breast Cancer Health Disparities Study

	Alcohol consumption						$P_{interaction}$
	Never			Ever			
	HR	95% CI	P-value	HR	95% CI	P-value	
rs1424482							0.362
T/T or T/C	1.00	-		1.00	-		
C/C	1.24	0.79 - 1.92	0.348	0.89	0.48 – 1.63	0.696	
rs168351							0.083
A/A	1.00	-		1.00	-		
A/G	1.12	0.73 - 1.73	0.598	0.61	0.35 - 1.05	0.075	
G/G	2.26	0.81 – 6.32	0.118	0.99	0.24 - 4.06	0.985	
rs1888202							0.497
C/C	1.00	-		1.00	-		
C/G or G/G	1.03	0.71 - 1.50	0.871	0.82	0.53 - 1.29	0.397	
rs348461							0.250
T/T	1.00	-		1.00	-		
T/A	1.23	0.86 - 1.76	0.260	0.75	0.49 - 1.14	0.173	
A/A	1.07	0.64 - 1.79	0.798	0.84	0.45 - 1.56	0.576	
rs348463							0.214
T/T	1.00	-		1.00	-		
T/C	1.32	0.94 - 1.87	0.113	0.81	0.53 - 1.24	0.332	
C/C	1.13	0.63 - 2.01	0.680	0.94	0.46 - 1.92	0.863	
rs348481							0.254
T/T	1.00	-		1.00	-		
T/C	1.01	0.70 - 1.47	0.938	0.76	0.49 - 1.20	0.238	
C/C	1.01	0.37 - 2.78	0.980	0.46	0.11 - 1.89	0.282	
rs63319							0.597
C/C or C/A	1.00	-		1.00	-		
A/A	1.06	0.72 - 1.56	0.757	1.28	0.83 - 1.97	0.271	
rs7027604							0.817
C/C or C/A	1.00	-		1.00	-		
A/A	1.18	0.80 - 1.73	0.396	1.28	0.81 - 2.04	0.290	
rs722921							0.898
T/T	1.00	-		1.00	-		
T/A or A/A	0.90	0.63 - 1.30	0.590	0.87	0.56 - 1.34	0.523	

Table 10. Associations between *ALDH1A1* polymorphisms and breast cancer-specific mortality, stratified by Native American (NA) ancestry, The Breast Cancer Health Disparities Study

	Percentage (%) of NA Ancestry						$P_{interaction}$
	% NA Ancestry ≤ 0.28¶			% NA Ancestry > 0.28			
	HR	**95% CI**	**P-value**	**HR**	**95% CI**	**P-value**	
rs1424482							0.673
T/T or T/C	1.00	-		1.00	-		
C/C	1.18	0.73 - 1.92	0.490	0.98	0.58 - 1.66	0.951	
rs168351							0.687
A/A	1.00	-		1.00	-		
A/G	0.87	0.58 - 1.29	0.483	0.88	0.47 - 1.65	0.691	
G/G	1.81	0.79 – 4.14	0.163	-	-	-	
rs1888202							0.778
C/C	1.00	-		1.00	-		
C/G or G/G	0.92	0.64-1.34	0.680	1.00	0.64 - 1.57	0.994	
rs348461							0.869
T/T	1.00	-		1.00	-		
T/A	1.07	0.75 - 1.52	0.710	0.92	0.60 - 1.42	0.723	
A/A	0.96	0.55 - 1.69	0.898	0.97	0.55 - 1.71	0.903	
rs348463							0.487
T/T	1.00	-		1.00	-		
T/C	1.13	0.81 - 1.59	0.480	1.05	0.69 - 1.60	0.824	
C/C	1.26	0.66 - 2.38	0.481	0.88	0.47 - 1.64	0.678	
rs348481							0.185
T/T	1.00	-		1.00	-		
T/C	0.99	0.70-1.40	0.959	0.72	0.43 - 1.23	0.230	
C/C	0.95	0.41-2.18	0.904	-	-	-	
rs63319							0.460
C/C or C/A	1.00	-		1.00	-		
A/A	1.25	0.87 - 1.81	0.222	0.98	0.61 - 1.57	0.930	
rs7027604							0.600
C/C or C/A	1.00	-		1.00	-		
A/A	1.30	0.88 - 1.91	0.188	1.09	0.69 - 1.73	0.698	
rs722921							0.907
T/T	1.00	-		1.00	-		
T/A or A/A	0.91	0.63 - 1.30	0.594	0.87	0.57 - 1.35	0.546	

FIGURES

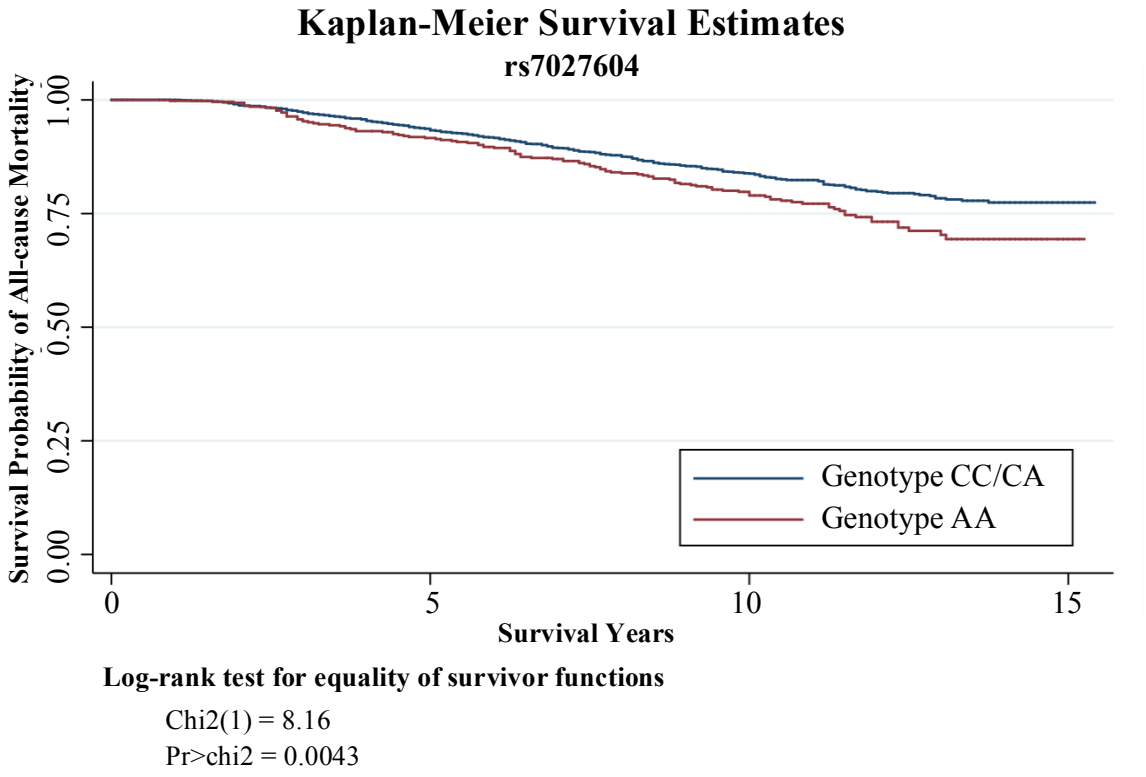

Figure 1. The Kaplan-Meier survival curves and log-rank test for genotype of the polymorphism rs7027604, taking into account the recessive mode of inheritance, The Breast Cancer Health Disparities Study

REFERENCES

1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2015.* CA Cancer J Clin, 2015. **65**(1): p. 5-29.
2. Yang, X. and M.E. Lippman, *BRCA1 and BRCA2 in breast cancer.* Breast cancer research and treatment, 1999. **54**(1): p. 1-10.
3. Clemons, M. and P. Goss, *Estrogen and the risk of breast cancer.* New England Journal of Medicine, 2001. **344**(4): p. 276-285.
4. Singletary, K.W. and S.M. Gapstur, *Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms.* Jama, 2001. **286**(17): p. 2143-2151.
5. Howe, G.R., et al., *Dietary factors and risk of breast cancer: combined analysis of 12 case—control studies.* Journal of the National Cancer Institute, 1990. **82**(7): p. 561-569.
6. White-Means, S., et al., *African American Women: Surviving Breast Cancer Mortality against the Highest Odds.* Int J Environ Res Public Health, 2015. **13**(1): p. ijerph13010006.
7. Reis, L., et al., *SEER cancer statistics review, 1973-1997.* Bethesda, MD: National Cancer Institute, 2000.
8. Newman, L.A., et al., *African-American ethnicity, socioeconomic status, and breast cancer survival: a meta-analysis of 14 studies involving over 10,000 African-American and 40,000 White American patients with carcinoma of the breast.* Cancer, 2002. **94**(11): p. 2844-54.
9. Chlebowski, R.T., et al., *Ethnicity and breast cancer: factors influencing differences in incidence and outcome.* J Natl Cancer Inst, 2005. **97**(6): p. 439-48.
10. Boone, S.D., et al., *The joint contribution of tumor phenotype and education to breast cancer survival disparity between Hispanic and non-Hispanic white women.* Cancer Causes Control, 2014. **25**(3): p. 273-82.
11. Vainshtein, J., *Disparities in breast cancer incidence across racial/ethnic strata and socioeconomic status: a systematic review.* Journal of the National Medical Association, 2008. **100**(7): p. 833-839.
12. Amey, C.H., M.K. Miller, and S.L. Albrecht, *The role of race and residence in determining stage at diagnosis of breast cancer.* The Journal of rural health, 1997. **13**(2): p. 99-108.
13. Ooi, S.L., M.E. Martinez, and C.I. Li, *Disparities in breast cancer characteristics and outcomes by race/ethnicity.* Breast cancer research and treatment, 2011. **127**(3): p. 729-738.
14. Chen, L. and C.I. Li, *Racial disparities in breast cancer diagnosis and treatment by hormone receptor and HER2 status.* Cancer Epidemiol Biomarkers Prev, 2015. **24**(11): p. 1666-72.
15. Roseland, M.E., et al., *Racial differences in breast cancer survival in a large urban integrated health system.* Cancer, 2015. **121**(20): p. 3668-75.

16. Parise, C.A. and V. Caggiano, *Disparities in race/ethnicity and socioeconomic status: risk of mortality of breast cancer patients in the California Cancer Registry, 2000–2010.* BMC cancer, 2013. **13**(1): p. 449.
17. Society, A.C., *Cancer facts and figures 2013*. 2013, American Cancer Society Atlanta.
18. Malone, K.E., et al., *Prevalence and predictors of BRCA1 and BRCA2 mutations in a population-based study of breast cancer in white and black American women ages 35 to 64 years.* Cancer Res, 2006. **66**(16): p. 8297-308.
19. Kurian, A.W., *BRCA1 and BRCA2 mutations across race and ethnicity: distribution and clinical implications.* Curr Opin Obstet Gynecol, 2010. **22**(1): p. 72-8.
20. Tian, N., et al., *Identifying risk factors for disparities in breast cancer mortality among African-American and Hispanic women.* Womens Health Issues, 2012. **22**(3): p. e267-76.
21. Paskett, E.D., *Breast Cancer Among Special Populations: Disparities in Care Across the Cancer Control Continuum.* Adv Exp Med Biol, 2015. **862**: p. 39-52.
22. Control, C.f.D. and Prevention, *Vital signs: racial disparities in breast cancer severity--United States, 2005-2009.* MMWR. Morbidity and mortality weekly report, 2012. **61**(45): p. 922.
23. Li, C.I., K.E. Malone, and J.R. Daling, *Differences in breast cancer stage, treatment, and survival by race and ethnicity.* Archives of internal medicine, 2003. **163**(1): p. 49-56.
24. Partridge, A.H., et al., *Subtype-Dependent Relationship Between Young Age at Diagnosis and Breast Cancer Survival.* J Clin Oncol, 2016. **34**(27): p. 3308-14.
25. Chen, H.L., et al., *Effect of Age on Breast Cancer Patient Prognoses: A Population-Based Study Using the SEER 18 Database.* PLoS One, 2016. **11**(10): p. e0165409.
26. Kelsey, J.L., M.D. Gammon, and E.M. John, *Reproductive factors and breast cancer.* Epidemiologic reviews, 1993. **15**(1): p. 36.
27. Connor, A.E., et al., *Pre-diagnostic breastfeeding, adiposity, and mortality among parous Hispanic and non-Hispanic white women with invasive breast cancer: the Breast Cancer Health Disparities Study.* Breast Cancer Res Treat, 2017. **161**(2): p. 321-331.
28. Islami, F., et al., *Breastfeeding and breast cancer risk by receptor status--a systematic review and meta-analysis.* Ann Oncol, 2015. **26**(12): p. 2398-407.
29. Sweeney, C., et al., *Reproductive history in relation to breast cancer risk among Hispanic and non-Hispanic white women.* Cancer Causes Control, 2008. **19**(4): p. 391-401.
30. Morimoto, L.M., et al., *Obesity, body size, and risk of postmenopausal breast cancer: the Women's Health Initiative (United States).* Cancer Causes & Control, 2002. **13**(8): p. 741-751.
31. Protani, M., M. Coory, and J.H. Martin, *Effect of obesity on survival of women with breast cancer: systematic review and meta-analysis.* Breast cancer research and treatment, 2010. **123**(3): p. 627-635.
32. Cancer, C.G.o.H.F.i.B., *Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with*

breast cancer and 101 986 women without the disease. The Lancet, 2001. **358**(9291): p. 1389-1399.
33. Sackey, H., et al., *The impact of in situ breast cancer and family history on risk of subsequent breast cancer events and mortality-a population-based study from Sweden.* Breast Cancer Research, 2016. **18**(1): p. 105.
34. Zakhari, S. and J.B. Hoek, *Alcohol and breast cancer: reconciling epidemiological and molecular data*, in *Biological Basis of Alcohol-Induced Cancer*. 2015, Springer. p. 7-39.
35. Longnecker, M.P., *Alcoholic beverage consumption in relation to risk of breast cancer: meta-analysis and review.* Cancer Causes and Control, 1994. **5**(1): p. 73-82.
36. Ellison, R.C., et al., *Exploring the relation of alcohol consumption to risk of breast cancer.* American Journal of Epidemiology, 2001. **154**(8): p. 740-747.
37. Rock, C.L. and W. Demark-Wahnefried, *Nutrition and survival after the diagnosis of breast cancer: a review of the evidence.* J Clin Oncol, 2002. **20**(15): p. 3302-16.
38. Thun, M.J., et al., *Alcohol consumption and mortality among middle-aged and elderly US adults.* New England Journal of Medicine, 1997. **337**(24): p. 1705-1714.
39. Smith-Warner, S.A., et al., *Alcohol and breast cancer in women: a pooled analysis of cohort studies.* Jama, 1998. **279**(7): p. 535-540.
40. Tomita, H., et al., *Aldehyde dehydrogenase 1A1 in stem cells and cancer.* Oncotarget, 2016. **7**(10): p. 11018-32.
41. Silva, I.A., et al., *Aldehyde dehydrogenase in combination with CD133 defines angiogenic ovarian cancer stem cells that portend poor patient survival.* Cancer research, 2011. **71**(11): p. 3991-4001.
42. Li, T., et al., *ALDH1A1 is a marker for malignant prostate stem cells and predictor of prostate cancer patients' outcome.* Laboratory investigation, 2010. **90**(2): p. 234-244.
43. Patel, M., et al., *ALDH1A1 and ALDH3A1 expression in lung cancers: correlation with histologic type and potential precursors.* Lung cancer, 2008. **59**(3): p. 340-349.
44. Kahlert, C., et al., *Low expression of aldehyde deyhdrogenase 1A1 (ALDH1A1) is a prognostic marker for poor survival in pancreatic cancer.* BMC cancer, 2011. **11**(1): p. 275.
45. Ginestier, C., et al., *ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome.* Cell stem cell, 2007. **1**(5): p. 555-567.
46. Zhong, Y., et al., *Expression of ALDH1 in breast invasive ductal carcinoma: an independent predictor of early tumor relapse.* Cancer cell international, 2013. **13**(1): p. 60.
47. Zhong, Y., et al., *ALDH1 is a better clinical indicator for relapse of invasive ductal breast cancer than the CD44+/CD24− phenotype.* Medical Oncology, 2014. **31**(3): p. 1-8.

48. Charafe-Jauffret, E., et al., *Aldehyde dehydrogenase 1–Positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer.* Clinical cancer research, 2010. **16**(1): p. 45-55.
49. Yao, S., et al., *Germline genetic variants in ABCB1, ABCC1 and ALDH1A1, and risk of hematological and gastrointestinal toxicities in a SWOG Phase III trial S0221 for breast cancer.* Pharmacogenomics J, 2014. **14**(3): p. 241-7.
50. Marcato, P., et al., *Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis.* Stem cells, 2011. **29**(1): p. 32-45.
51. Tan, E., A. Thike, and P. Tan, *ALDH1 expression is enriched in breast cancers arising in young women but does not predict outcome.* British journal of cancer, 2013. **109**(1): p. 109-113.
52. Resetkova, E., et al., *Prognostic impact of ALDH1 in breast cancer: a story of stem cells and tumor microenvironment.* Breast cancer research and treatment, 2010. **123**(1): p. 97-108.
53. Spence, J.P., et al., *Evaluation of aldehyde dehydrogenase 1 promoter polymorphisms identified in human populations.* Alcoholism: Clinical and Experimental Research, 2003. **27**(9): p. 1389-1394.
54. Yoshida, A., et al., *Cytosolic aldehyde dehydrogenase (ALDH1) variants found in alcohol flushers.* Annals of human genetics, 1989. **53**(1): p. 1-7.
55. Quertemont, E., *Genetic polymorphism in ethanol metabolism: acetaldehyde contribution to alcohol abuse and alcoholism*. 2004, Nature Publishing Group.
56. Ehlers, C.L., et al., *Association of ALDH1 Promoter Polymorphisms With Alcohol‐Related Phenotypes in Southwest California Indians.* Alcoholism: Clinical and Experimental Research, 2004. **28**(10): p. 1481-1486.
57. Moore, S., et al., *Association of ALDH1 promoter polymorphisms with alcohol-related phenotypes in Trinidad and Tobago.* Journal of studies on alcohol and drugs, 2007. **68**(2): p. 192-196.
58. Crawford, A., et al., *Haplotype-based study of the association of alcohol and acetaldehyde-metabolising genes with alcohol dependence (with or without comorbid anxiety symptoms) in a Cape Mixed Ancestry population.* Metab Brain Dis, 2014. **29**(2): p. 333-40.
59. Slattery, M.L., et al., *Genetic variation in genes involved in hormones, inflammation and energetic factors and breast cancer risk in an admixed population.* Carcinogenesis, 2012. **33**(8): p. 1512-21.
60. Slattery, M.L., et al., *Body size, weight change, fat distribution and breast cancer risk in Hispanic and non-Hispanic white women.* Breast cancer research and treatment, 2007. **102**(1): p. 85-101.
61. Connor, A.E., et al., *Ethnic differences in the relationships between diabetes, early age adiposity and mortality among breast cancer survivors: the Breast Cancer Health Disparities Study.* Breast Cancer Res Treat, 2016. **157**(1): p. 167-78.
62. Connor, A.E., et al., *Associations between TCF7L2 polymorphisms and risk of breast cancer among Hispanic and non-Hispanic white women: the Breast*

Cancer Health Disparities Study. Breast cancer research and treatment, 2012. **136**(2): p. 593-602.
63. Connor, A.E., et al., *Associations between ALOX, COX, and CRP polymorphisms and breast cancer among Hispanic and non‐Hispanic white women: The breast cancer health disparities study.* Molecular carcinogenesis, 2015. **54**(12): p. 1541-1553.
64. Falush, D., M. Stephens, and J.K. Pritchard, *Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies.* Genetics, 2003. **164**(4): p. 1567-1587.
65. Pritchard, J.K., M. Stephens, and P. Donnelly, *Inference of population structure using multilocus genotype data.* Genetics, 2000. **155**(2): p. 945-959.
66. Hosmer, D., *V, Lerneshow S: Applied Logistic Regression*. 1989, New York, NY, John Wiley and Sons.
67. Dupont, W.D. and W.D. Plummer, *Power and sample size calculations: a review and computer program.* Controlled clinical trials, 1990. **11**(2): p. 116-128.
68. Owzar, K., et al., *Power and Sample Size Calculations for SNP Association Studies With Censored Time‐to‐Event Outcomes.* Genetic epidemiology, 2012. **36**(6): p. 538-548.
69. Schwartz, T., et al., *Expression of aldehyde dehydrogenase 1 as a marker of mammary stem cells in benign and malignant breast lesions of Ghanaian women.* Cancer, 2013. **119**(3): p. 488-494.
70. Liu, J., et al., *Haplotype-based study of the association of alcohol-metabolizing genes with alcohol dependence in four independent populations.* Alcohol Clin Exp Res, 2011. **35**(2): p. 304-16.

CURRICULUM VITAE

PERSONAL INFORMATION

Name: Zhiyu Xia
Date of birth: January 22nd 1992
Place of birth: Chendu, Sichuan Province, China
Mobile: 410-972-9300
Email: zhiyuxia7@gmail.com

EDUCATION

Master of Science in Cancer Epidemiology
Johns Hopkins Bloomberg School of Public Health (Baltimore, MD, USA)
08/2015 – 05/2017
GPA: 3.95/4.00 (up to current term)
Thesis: *"Associations between ALDH1A1 polymorphisms, alcohol consumption, and mortality among Hispanic and non-Hispanic white women diagnosed with breast cancer: The Breast Cancer Health Disparities Study (BCHDS)"*
Advisor: Dr. Avonne Connor

Bachelor of Medicine in Public Health
Peking University Health Science Center (Beijing, China)
09/2010 – 07/2015
GPA: 3.36/4.00 (major GPA: 3.55/4.00)
Thesis: *"Comparison of Reproductive Health between Migrant Youths and Those Living In Rural Areas"*
Advisor: Dr. Ying Ji

Exchange student in the Global Learning Semester Program
Duke Kunshan University (Kunshan, China)
08/2014 – 12/2014
GPA: 3.68/4.00

RESEARCH EXPERIENCE

Research Assistant, Supervisor: Dr. Pamela Surkan and Dr. Alison Abraham
Johns Hopkins Bloomberg School of Public Health
06/2016 – 05/2017
Cultural and biological susceptibility to depression: explaining racial differences in reported symptoms of depression
- Examined the concordance of self-perceived and biological race in men who have sex with men (MSM), using Multicenter AIDS Cohort Study (MACS)
- Identified potential risk and protective factors in relation to racial disparities in depressive symptoms among MSM using Generalized Estimating Equation (GEE) and mixed-effect models

The rate of decline in grip strength among HIV+ and HIV- adults: Multicenter AIDS Cohort Study (MACS)
- Evaluated and compared the rates of decline in grip strength among HIV+ and HIV- adults stratified by age groups

- Conducted longitudinal analysis using mixed-effect models with empirical Bayes method. Joint models combining mixed-effect and time-to-event were applied to address informative loss to follow-up

Research Assistant, Supervisor: Dr. Alison Klein
Johns Hopkins School of Medicine
01/2016 – 12/2016

Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer: National Familial Pancreas Tumor Registry (NFPTR)

- Replicated Genome-wide Association Studies (GWAS) in Plink using genetic data from Pancreatic Cancer Case-Control Consortium (PanC4), meta-analysis of the PanC4 data with data from PanScan 1 and PanScan 2, and imputation using HapMap reference panels

Research Assistant, Supervisor: Dr. Tonia Poteat
Johns Hopkins Bloomberg School of Public Health
10/2015 – 07/2016

Global epidemiology of HIV infection and related syndemics in transgender people

- Conducted systematic review on the prevalence of HIV and co-infected STIs in transgender people

Research Assistant, Supervisor: Dr. Justin Lessler
Johns Hopkins Bloomberg School of Public Health
11/2015 – 02/2017

JHSPH Global Cholera project

- Conducted data extraction and entry for the surveillance data

Research Assistant, Supervisor: Dr. Jianping Chen
University of Hong Kong School of Chinese Medicine
07/2014 – 02/2016

- Conducted a case-control study to investigate the relationship between dietary patterns of Traditional Chinese Medicine (TCM) and the risk of breast cancer among Hong Kong Chinese women at Mary Hospital
- Evaluated the risk factors of left-sided predominance among Hong Kong Chinese women with breast cancer
- Estimated the efficacy of Guolin-Qigong on body-mind health of Chinese women with breast cancer in a randomized controlled trial (RCT)

Research Assistant, Supervisor: Dr. Kit Chan
University of Birmingham
03/2015 – 08/2015

Systematic Review of the Prevalence of Alzheimer's Disease in China During 2011 and 2015

- Conducted systematic review on the prevalence of subtypes of Alzheimer's disease in different geographic areas of China during 2011 and 2015 by synthesizing evidence from various study designs

Research Assistant, Supervisor: Dr. Yonghua Hu
Peking University Department of Epidemiology
07/2014 – 08/2014

Family Cohort Study of Chronic Non-communicable Diseases in Northern Rural Areas in China, sponsored by National Natural Science Foundation of China

- Collected data from participants via in-person interviews from 4 villages in rural Beijing areas

- Conducted fieldwork by doing ECG measurement, collecting tissue samples, and preparing blood samples

Research Assistant, Supervisor: Dr. Xiaochuan Pan
Peking University Department of Occupational and Environmental Health
03/2014 – 07/2014
Comparative Study for the Different Air Temperature Indicators and Respiratory Mortality of the Population

- Extracted data from medical records and questionnaires on daily mortality due to respiratory diseases, air temperature, air pollutants, and other meteorological factors in Haidian district between 2004 and 2008

TEACHING EXPERIENCE

Teaching Assistant, Instructors: Dr. David Dowdy and Dr. Keri Althoff
Johns Hopkins Bloomberg School of Public Health
01/2017 – 03/2017
340.753.01 EPIDEMIOLOGIC METHODS 3 (5 credits), 3rd term

INTERNSHIP AND OTHER PROFESSIONAL ACTIVITIES

Intern, Beijing Office of "Advanced Breast Cancer Awareness in China"
07/2013 – 08/2013

Intern, Beijing Haidian Center for Disease Control and Prevention
07/2011& 07/2014

Clinical Intern, Beijing Shijitan Hospital
02/2013 – 01/2014

International Medical Volunteer, Team of PKU Sunshine & Love Clinic visiting the U.S.
01/2013 – 02/2013

Volunteer, Peking University Sunshine & Love Clinic
09/2011 – 07/2015

Team Leader, Community Immersion Team of Peking University Visiting Yunnan Province
07/2012 – 08/2012

Vice-president, Student Assembly of School of Public Health, Peking University
10/2011 – 12/2012

Officer, DKU International Association, Duke Kunshan University
09/2014 – 12/2014

Volunteer (selected):
5th International Hot Topics Heart Forum and Cardiac Inter-discipline Forum (2013)
1st Aid For Health Simulation, Peking University (2014)
3rd Global Health Institutional Partnership Network Meeting, DKU (2014)
Office of Youth League Committee, Peking University (2010-2011)
Loveheart Organization, Peking University (2010-2011)

PUBLICATIONS

Published and Accepted

Zheng, X., Xie, T., **Xia, Z.**, Chen, J. An Epidemiological Study of the Relationship between Dietary Patterns of TCM and the Risk of Breast Cancer in Hong Kong Chinese Women. Chinese Medicine (accepted)

Liang, F., Xu, M., Jin, X., Li, G., Tian, L., **Xia, Z.**, Pan, X. (2014). Comparative study for the different air temperature indicators and respiratory mortality of the population. Journal of Environment and Health, 5, 001

Submitted

Zheng, X., Diu, C., **Xia, Z.**, Chen, J. Predominance of left-sided breast cancer: an explanation from a traditional Chinese medicine approach and its treatment implication. Journal of Translational Medicine (submitted)

Liu, P., You, J., **Xia, Z.**, Chen, J. The efficacy of Guolin-Qigong on body-mind health of Chinese women with breast cancer: a randomized controlled trial. Plos One (submitted)

SCHOLARLY PRESENTATIONS

Associations between ALDH1A1 polymorphisms, alcohol consumption, and mortality among Hispanic and non-Hispanic white women diagnosed with breast cancer: The Breast Cancer Health Disparities Study. Poster presentation. American Association for Cancer Research (AACR) 2017 Annual Meeting, Washington DC, USA

The effect of obesity on the incidence of depression in the participants of Baltimore Memory Study (BMS). Poster presentation. Methodologic Challenges in Epidemiologic Research (Spring 2016) class final project showcase, Baltimore, MD, USA

COMPUTER SKILLS

Stata, SAS, R, SPSS, Plink, Merlin, Epidata, Endnote, Microsoft Office Software, research engines